P-ISSN: 2618-0723
E-ISSN: 2618-0731

NAAS Rating (2025): 5.04
www.extensionjournal.com

International Journal of Agriculture Extension and Social Development

Volume 8; Issue 9; September 2025; Page No. 690-698

Received: 03-07-2025
Accepted: 10-08-2025

Indexed Journal
Peer Reviewed Journal

A hybrid machine learning model for crop yield prediction based on meteorological data and pesticide information

[1]Chandra Sekhar Sanaboina and [2]K Vidya Sree

[1]Professor, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India

[2]M.Tech, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India

DOI: https://www.doi.org/10.33545/26180723.2025.v8.i9j.2498

Corresponding Author: Chandra Sekhar Sanaboina

Abstract

This study article's main goal is to provide a reliable method of agricultural production prediction by utilizing historical crop yield data, pesticide usage, and climatic data. To make precise predictions, the proposed model has a hybrid approach and uses advanced machine learning algorithms like Random Forest Regression, Extreme Gradient Boosting (XGBoost) and K-Nearest Neighbors (KNN). The paper also discusses ensemble methods such as Random Forest and XGBoost ensemble methods or XGBoost and KNN ensemble methods to achieve predictive power and model stability. The model is trained and tested on comprehensive datasets of solar radiation, meteorological records and Indian agriculture and climate data. Key methodologies used include K-Fold cross validation for hyper-parameter tuning using GridSearchCV, feature selection methods to make the model more focused, and ensemble learning to make the model robust and reduce bias. The entire system is created using Python, making use of powerful libraries such as Sickie-learn and Tensor flow. Achieved a high predictive accuracy, and XGBoost surpassed other models by 95% Feature selection consisted of meteorological variables, pesticide use and crop yield records, which improved model performance significantly. These findings show that the suggested hybrid strategy for crop production prediction is reliable.

Keywords: K-Fold cross-validation, GridSearchCV, feature selection, ensemble methods, Xtreme Gradient Boosting (XGBoost), Random Forest regression, Soil Moisture and Nutrient Levels, Solar Radiation Data, Meteorological Dataset

1. Introduction

Agriculture is a vital sector for the economy of any country, especially developing countries such as India, where a large part of the population depends on agriculture for their livelihood. Food security, economic growth, and other aspects of rural development depend on consistent and predictable agricultural productivity. However, because agricultural and environmental elements are dynamic and frequently unpredictable, forecasting crop yields is extremely difficult. Climate change has had a greater impact on traditional farming practices, making it more challenging for farmers to rely only on past performance due to unpredictable rainfall patterns, rising temperatures, and extreme weather events.

The abuse and improper management of agricultural inputs like pesticides and fertilizers has added to the complexities. While these are designed to protect crops, when misused can bring degradation of soil, damage to beneficial organisms and ultimately, reduced productivity. This situation brings into focus the need for more sophisticated farming systems which would be able to analyse the combined effects of weather patterns and chemical inputs to provide more informed and sustainable farming decisions. The move towards data-driven agriculture, the availability of large-scale datasets and advances in data processing has paved the way for new opportunities in precision farming and better controlling and predicting crop yields.

Machine learning (ML) is revolutionary technologies that have the power to make a significant change. As a branch of artificial intelligence it is based on the use of algorithms that help to identify patterns in data and draw predictions without having to explicitly program for every possible case. This research article uses the power of ML to create a better and more comprehensive model of predicting the crop yield. By combining meteorological data with data on pesticides and crop harvests, the system hopes to provide farmers with valuable insights to optimise their planting and harvest strategies and improve both productivity and sustainability in the agricultural sector.

1.1 Random Forest Regression: A strong and adaptable supervised learning approach for classification and regression applications is the Random Forest algorithm.

It belongs to the family of ensemble learning techniques, more specifically of bagging (Bootstrap Aggregating). The algorithm generates multiple decision trees as a part of the training procedure and produce the final decision by computing the majority decision in the case of classification problem or the average value in the case of regression problem from all of the trees.

1.2 XGBoost: XGBoost or Extreme Gradient Boosting, is a high-performance machine learning algorithm which is widely used for classification and regression problems. It creates models one after the other in such a way that each new model is concerned with how to fix the errors in the previous model. This process is grounded on gradient-boosting, which operates by a process called gradient descent, a method for minimizing a loss function, in order to make a better prediction. XGBoost is unique in that it is designed to be fast, scalable and efficient, which makes it a popular choice for large datasets.

One of the strengths of XGBoost is the regularization that comes with it that includes L1 and L2 methods in order to reduce over fitting and generalization. It has in-built handling of missing data as well so that the algorithm works fine even when some features are missing. Additionally, XGBoost supports parallel processing which makes it faster to train than traditional gradient boosting methods. These enhancements make it more robust and reliable at the variety of tasks.

1.3 K-Nearest Neighbors: K-Nearest Neighbors, or KNN, is a straightforward, instance-based, non-parametric learning technique that works well for both regression and classification. KNN uses its 'k' nearest neighbors to forecast a new data point based on the majority class (classification) or average value (regression) in the training data. The 'k' is an integer that is (user-defined). Metrics like the Euclidean distance are commonly used to compute the distance between data points.

1.3.1 'k' Parameter: The number of closest neighbors that will be taken into account when making a forecast. Choosing the value of 'k' is crucial since it has a significant impact on the model's performance. Simple to understand and implement, no training phase (or minimal) can handle multi-class problems. The method is sensitive to feature scaling, it may be difficult to choose the suitable value of 'k' and to select the suitable distance metric.

Literature Review

Recent developments in agriculture have focused on techniques to combine the use of satellite data, remote sensing, and machine learning capabilities to enable the accurate monitoring and analysis of crops. Spatiotemporal image fusion methods have enhanced the resolution for crop monitoring at subfield-level [1]. Gene-related studies in crops like maize have demonstrated how omics information may be combined with computational methods for yield-related information [2]. At the same time, this has enabled that the usage of interpretable deep learning models like LSTM networks have been made possible, for both estimating crop yield and keeping the prediction transparent [3]. Soil and environmental factors have also been integrated into land suitability models using feature selection and classification techniques [4] which makes the predictions more reliable.

Machine learning has also been applied to monitoring the quality of the crops and yield estimation. Predictions models in real time have been developed for crops like soybean during transportation to preserve quality [5]. High resolution data collection using UAV-based LiDAR and multi-spectral imaging has been proven to be used for increasing crop trait estimation [6]. Further improvement in the accuracy of crop yield prediction has been obtained by advanced neural architectures such as ConvLSTM and hybrid models using vision transformer with Earth observation datasets [7]. In addition, crop recommendation systems as well as planning frameworks based on intelligent algorithms have come up to support farmers in choosing appropriate crop as well as optimizing agricultural planning [8].

Classification techniques with radar and time series satellite data have been helpful in differentiating between different types of crops [9]. Neural networks with attention mechanisms and 3D-CNN architectures have also been attempted to further enhance multispectral yield prediction models [10]. Semantic segmentation techniques have been applied for multitemporal datasets for better recognition of crops [11]. Recently, approaches based on reinforcement learning techniques were applied to farm level crop planning to support sustainable agricultural practices [12]. These developments show how deep learning and systems based on artificial intelligence move away from traditional forecasting and more towards adaptive and prescriptive systems.

Another important trend is that of explainable artificial intelligence in agriculture. Reliability scores from saliency-based approaches have been used for predicting harvest readiness and explain ability have been also highlighted in subfield level yield predictions [13]. Intelligent IoT devices have been developed for designing disease detection, irrigation management, and crop selection in a single system that supports smart agriculture practises [14]. New prediction models have been presented to enhance the productivity of crops by learning frameworks that focus on yield [15]. Weather-based prediction systems have also been developed at broader scales and can provide information for food security and climate adaptation [16].

Recent research has focused on better classifying and predicting crop yield using powerful and hybrid machine learning methods. These models solve the problems of scalability, data variability and sustainability in agriculture [17]. Novel transformer-based models and graph-based models have shown promising results in multimodal crop yield prediction [18], [19].Remote sensing data has been further applied for food security-focused yield estimation, while IoT-enabled models provide weekly pest forecasts for farmers [20], [23]. Efforts to integrate deep learning with explainable approaches have also been reported in plant disease detection [24].

Beyond crop yield, research has extended into resource optimization and multi-objective agricultural systems. Agrivoltaic models, for example, optimize solar energy production while ensuring crops receive adequate light [25]. Ensemble learning and hybrid models have been applied for crops such as sugarcane, rice, and cotton to achieve higher yield predictability [26], [29]. New adaptive mixture models improve area estimation in crop monitoring, while genetic algorithms enhance prediction efficiency when combined with machine learning [31], [32]. Additionally, irrigation-level prediction systems and semantic segmentation for major crops support smart farm management practices [33].

Overall, the literature highlights a transition toward intelligent, explainable, and scalable systems in agriculture. Advanced neural models, IoT integration, and reinforcement learning approaches are enabling not just prediction but also decision support at both farm and regional levels. The focus is increasingly on generalization across regions, sustainability-driven predictions, and multi-objective optimization to balance yield, quality, and resource use [34], [35]. These developments demonstrate that future agricultural systems will rely heavily on integrated AI and remote sensing technologies for sustainable food production.

System Architecture

Figure 1 depicts the architecture and overall workflow of the proposed crop yield prediction based on meteorological data

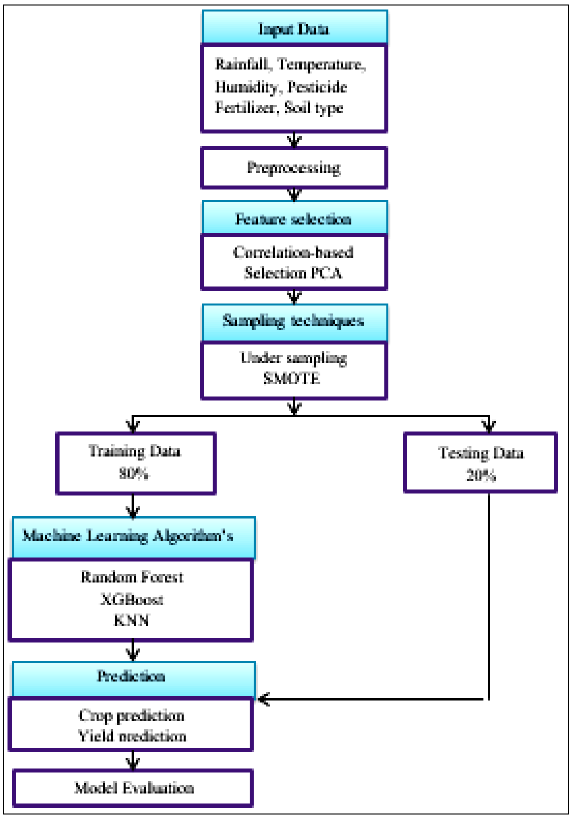

Fig 1: Architecture of the Proposed System

3.1 Crop Dataset

This section contains the datasets to train and test the prediction model for crop. It consists of data relating to meteorological parameters (e.g. rainfall, temperature, humidity), pesticide use, fertilizer application and historical crop yield statistics. These data sets are typically collected from public databases on agriculture or meteorological services. The dataset is the foundation of the model, and it provides the input features and target variables that are required to learn patterns and relationships.

3.2 Data Pre-processing

The process of cleaning and preparing raw data for machine learning models is known as data pre-processing. This can include dealing with missing values, encoding categorical data with label encoding, and normalizing numerical features with methods such as Min-Max Scaling. These include important steps to ensure the quality of the data, which can help to improve the accuracy of models and avoid issues such as over fitting or biased predictions.

3.3 Futures Selection and Sampling Method

This section is responsible for the identification of the most relevant features that have the biggest effect on the prediction of crop yield and of the crop type. Feature selection helps in reducing the dimensions, increasing the model performance and also increases the training time. Sampling techniques (such as under-sampling or SMote) may be used as an approach to deal with class imbalance, for example, when certain crops are more or less represented in the dataset.

3.4 Crop Prediction

In this stage, machine learning algorithms such as Random Forest, XG Boost, KNN and their hybrid combinations (RF +XG Boost, XG Boost +KNN) are used to predict the crop which is best suited in the given input conditions like temperature, humidity, pesticide level and soil features. This is useful for farmers or agricultural planners in their crop choices based on environmental and agricultural parameters.

3.5 Yield Prediction

Once the crop is predicted, the next step is to estimate the predicted yield in tons/hectare using regression models. This is done with same or extended feature set, and focus on variables most strongly correlated with output yield. The yield prediction helps in estimating productivity, planning storage, pricing, and agricultural supply chain logistic.

4. Methodology

4.1 Data

This stage involves collecting various datasets relevant to agriculture datasets are collected to build the foundation for training machine learning models. The data includes meteorological factors such as temperature, rainfall, humidity, and solar radiation, along with pesticide and fertilizer information covering their types and quantities used. Additionally, crop yield data is gathered, representing the historical production records of different crops. These datasets are obtained from open repositories, including platforms like Kaggle and government agricultural portals, ensuring reliable source for model development

Table 1: Dataset Details for Crop Yield and Type Prediction

Crop	Year	State	Pesticides	Yield	Temperature	Humidity
Barley	2019	Madhya Pradesh	11002.69	2.414286	28.83	56.35
Wheat	2016	AP	11.55	2.155	35.48	68.25
Maize	2019	Karnataka	501468.8	3.572069	28.7	50.97
Sugarcane	2021	Odisha	4504.3	67.76214	40	75.07
Tobacco	2017	Tamil Nadu	1439.44	1.594545	29.29	73.05

4.2 Training
Using the train test split function, the processed dataset—which includes X, crop yield, and the original crop labels—is divided into training and testing datasets. An 80/20 split is employed, in which the models are trained using 80% of the data and their performance is assessed using 20%.

Model Selection and Training: Three different machine learning algorithms are selected and trained:

4.2.1 Random Forest: Both a Random Forest classifier (for crop prediction) and a Random Forest regressor (for the prediction of yield) are trained. GridSearchCV is used for optimizing the hyper parameters i.e. n_estimators and max_depth. GridSearchCV is used to optimise the hyper parameters, n_estimators and max_depth.

4.2.2 XGBoost: An XGBClassifier and an XGBRegressor are trained using GridSearchCV for the hyperparameters (n_estimators) as well.

4.2.3 KNN: A K Neighbors Classifier and K Neighbors Regressor is trained with the help of hyperparameters (n neibors) using GridSearchCV

4.3 Evaluation
The performance of the models were evaluated separately and in combination. For the prediction of crops, classifiers (Random Forest, XGBoost and KNN) were evaluated using accuracy scores and for yield prediction, mean absolute error (MAE) scores were used. In addition, ensemble models using Random Forest with XGBoost and XGBoost with KNN were also tested, where the same evaluation metrics were used. This comparison highlights both the effectiveness of individual models and the improvements gained through ensemble learning.

4.5 Comparison: The calculated metrics for individual models and ensemble models are compared to assess the effectiveness of the ensemble approaches. Visualizations (bar charts) are used to facilitate this comparison.

Results and performance analysis

Table 2: Model Accuracy Results for Crop Prediction

	Algorithm Name	Accuracy
1	KNN	50.2%
2	Random Forest	76.6%
3	XGBoost	95.4%

Table 2 the results, it is clear that XGBoost achieved the highest accuracy of 0.954, showing strong predictive performance. In comparison, Random Forest reached 0.766 accuracy, while KNN performed the lowest with 0.502.

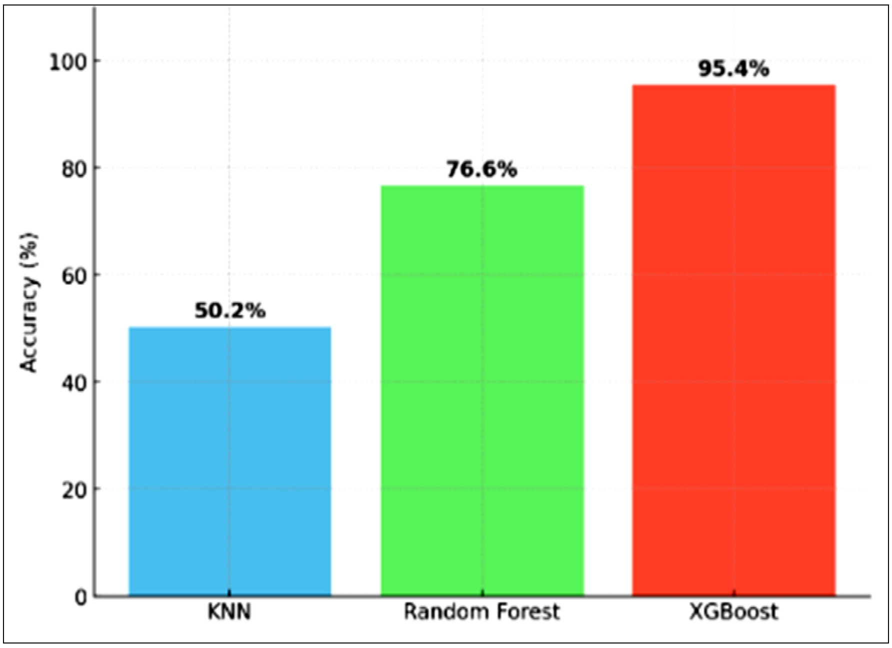

Fig 2: Accuracy Comparison of Models in Crop Prediction

Figure 2 shows the accuracy performance of three machine learning models used for crop yield prediction. Among them, XGBoostachieved the highest accuracy of 0.95, showing strong predictive capability. XGBoost attained the top accuracy of 0.95, indicating its excellent ability to make accurate predictions. Random Forest followed with a decent accuracy of 0.77, while KNN performed the lowest at 0.50, indicating weaker generalization on the dataset. This comparison highlights that XGBoost is the most effective model for this task.

Table 3: Performance Analysis of Regression Models in Yield Prediction

Model	MAE	RMSE	R^2 Score
KNN	0.0191	0.0495	0.488
Random Forest	0.0177	0.0384	0.758
XGBoost	0.0199	0.0337	0.954

Table 3 Based on the evaluation metrics, XGBoost achieved the best performance with an R^2 score of 0.954 and the lowest RMSE of 0.0337, proving its strong predictive

ability. Random Forest showed moderate accuracy with an R^2 of 0.758, while KNN performed the weakest with only 0.488.

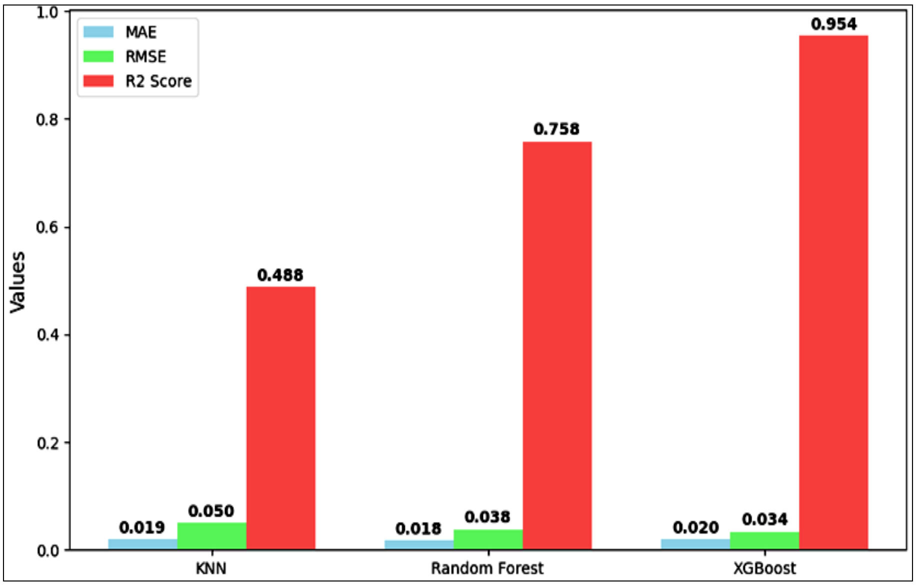

Fig 3: Performance Metrics for Yield Prediction

Three machine learning models—Random Forest, XGBoost, and KNN—are compared in Figure 3 using MAE, RMSE, and R^2 Score. With the lowest error numbers and the greatest R^2 Score among them, XGBoost performs better than the others, demonstrating its excellent prediction accuracy. Random Forest also performs reasonably well, though slightly less accurate than XGBoost. In contrast, KNN shows the highest error rates and the lowest R^2 Score, suggesting it is less effective for this prediction task. This highlights XGBoost as the most reliable model among the three for crop yield prediction.

Table 4: Regression Performance Analysis of Random Forest + KNN (Hybrid Model)

Model	MAE	RMSE	R^2 Score
Random Forest	2.63	3.24	0.770
KNN	3.97	4.45	0.488
RF + KNN	3.30	3.84	0.629

Table 4 performance comparison between Random Forest, KNN, and their hybrid model (RF + KNN) shows that Random Forest achieved better accuracy with lower MAE (2.63) and RMSE (3.24) along with a higher R^2 score of 0.770. KNN performed relatively weaker with higher error values and an R^2 of 0.488. The combined RF + KNN model provided balanced results with MAE of 3.30, RMSE of 3.84, and R^2 of 0.629, indicating that the hybrid approach improves stability compared to KNN while maintaining competitive accuracy.

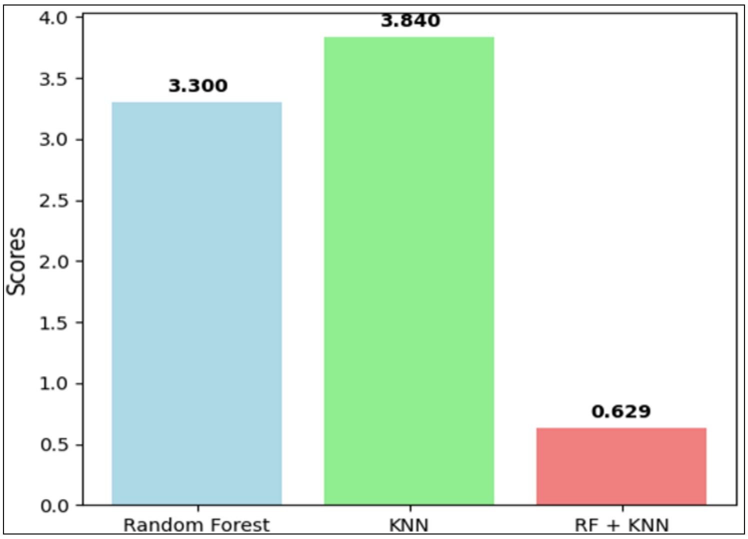

Fig 4: Accuracy Comparison of ML Models and RF + KNN (Ensemble)

Figure 4 represents the evaluation of the RF + KNN hybrid model using three key performance metrics: MAE, RMSE, and R² score. It highlights that the model achieves a balanced outcome, with acceptable error levels and a reasonable ability to explain the data variability. This indicates that combining Random Forest and KNN enhances the overall stability and performance compared to using the individual models.

Table 5: Hybrid RF + KNN Model with PCA and SMOTE for Yield Prediction

Model Variant	MAE	RMSE	R^2 Score
Random Forest	2.63	3.24	0.770
KNN	3.97	4.45	0.488
RF + KNN	3.30	3.84	0.629
RF + KNN + PCA	3.12	3.70	0.648
RF + KNN + SMOTE	2.95	3.55	0.672
RF + KNN + PCA + SMOTE	2.81	3.40	0.695

Table 5 experimental results show that combining Random Forest with KNN improves performance compared to individual models. Further enhancements using PCA and SMOTE gradually reduced MAE and RMSE while increasing the R² score, demonstrating better accuracy and robustness in prediction.

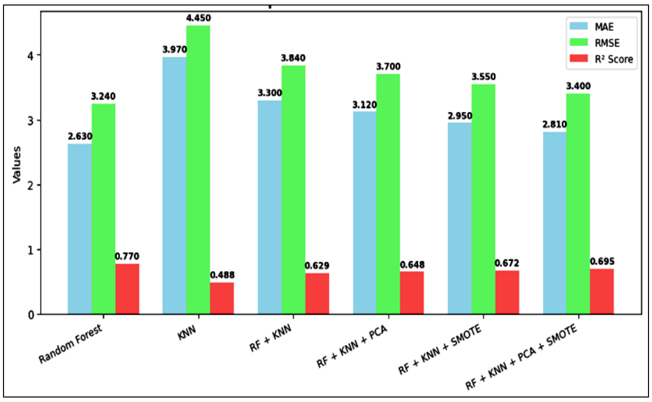

Fig 5: Performance Comparison of RF + KNN Variants with PCA and SMOTE

Figure 5 comparison of Random Forest, KNN, and their hybrid models with PCA and SMOTE. Results indicate that MAE and RMSE decrease while R² improves, with the RF + KNN + PCA + SMOTE model giving the best performance.

Table 6: Performance Analysis of Random Forest + XGBoost (Hybrid Model)

Model	MAE	RMSE	R^2 Score
Random Forest	2.63	3.24	0.770
XGBoost	1.15	1.87	0.950
RF + XGBoost	2.31	2.91	0.816

Table 6 comparison, XGBoost outperformed the other models with the lowest MAE (1.15), RMSE (1.87), and the highest R² score of 0.950, indicating strong prediction accuracy. The hybrid RF +XGBoost model performed better than Random Forest alone, but still could not surpass XGBoost.

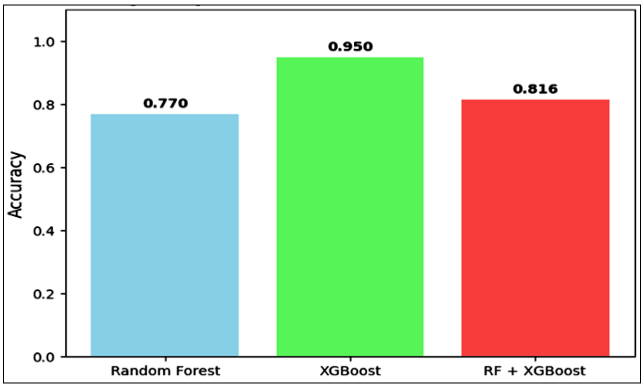

Fig 6: Accuracy Comparison of ML Models and RF + XGBoost (Ensemble)

Figure 6 combination of Random Forest and XGBoost enhances model performance by blending the strengths of bagging and boosting techniques. Random Forest captures general patterns, while XGBoost refines errors through gradient boosting. This hybrid reduces MAE and RMSE and improves R², resulting in more accurate crop yield predictions.

Table 7: Hybrid RF +XGBoost Model with PCA and SMOTE for Yield Prediction

Model Variant	MAE	RMSE	R^2 Score
RF + XGBoost(Original)	2.31	2.91	0.816
RF + XGBoost + PCA	2.18	2.80	0.828
RF + XGBoost + SMOTE	2.04	2.65	0.845
RF + XGBoost + PCA + SMOTE	1.92	2.51	0.861

Table 7 after applying PCA, the RF +XGBoost model showed improved performance by reducing redundant features and focusing on the most important components, which lowered error values slightly. When SMOTE was used, the model achieved even better accuracy as the balanced dataset helped in reducing bias and improving prediction stability. Overall, SMOTE provided the most significant improvement, giving lower MAE and RMSE with a higher R² score.

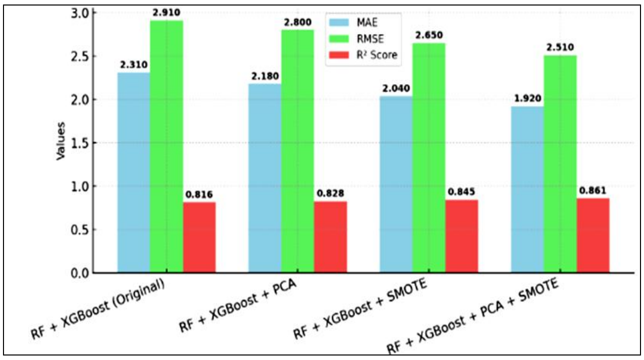

Fig 7: Performance Comparison of RF + XGBoost Variants with PCA and SMOTE

Figure 7 compares the performance of RF + XGBoost models with different enhancements such as PCA and SMOTE. It shows a steady reduction in MAE and RMSE values, along with an improvement in the R² score when these techniques are applied. The RF + XGBoost + PCA + SMOTE variant performs best, indicating its effectiveness in achieving higher accuracy.

Table 8: Regression Performance Analysis of XGBoost + KNN (Hybrid Model)

Model	MAE	RMSE	R^2 Score
XGBoost	1.15	1.87	0.950
KNN	3.97	4.45	0.488
XGBoost +KNN	2.89	3.20	0.710

Table 8 results, XGBoost achieved the best performance with the lowest error values (MAE 1.15, RMSE 1.87) and the highest R^2 score of 0.950. The hybrid XGBoost +KNN model showed improved results compared to KNN alone, but it still lagged behind the standalone XGBoost

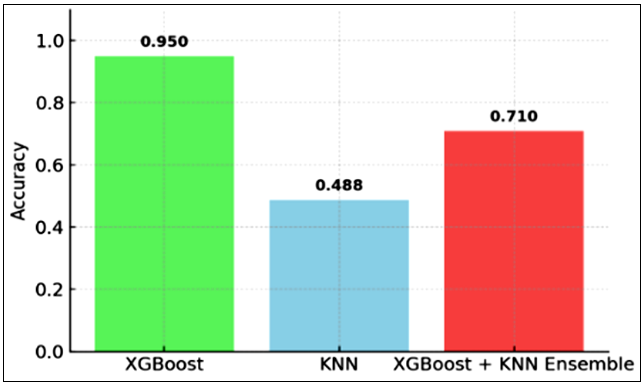

Fig 8: Accuracy Comparison of ML Models and XGBoost + KNN

Figure 8Merging XGBoost with KNN allows the model to learn global trends and refine local variations. XGBoost first models the overall structure, and KNN adjusts predictions based on nearby data points. This combination achieves better accuracy, reduced error rates, and stronger R^2 compared to standalone models.

Table 9: Hybrid XGBoost +KNN Model with PCA and SMOTE for Yield Prediction

Model Variant	MAE	RMSE	R^2 Score
XGBoost + KNN(Original)	2.89	3.20	0.710
XGBoost + KNN + PCA	2.74	3.05	0.732
XGBoost + KNN + SMOTE	2.61	2.91	0.756
XGBoost + KNN + PCA + SMOTE	2.47	2.78	0.773

Table 9 hybrid XGBoost + KNN model showed improved results after applying PCA, as dimensionality reduction helped to remove redundant information and focus on essential features, leading to lower error values. With SMOTE, the model achieved even better performance since balancing the dataset reduced bias and enhanced prediction accuracy. Overall, SMOTE provided the highest improvement with the lowest MAE and RMSE along with a better R^2 score.

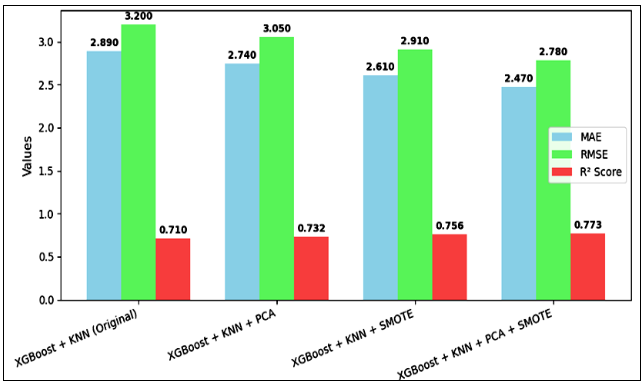

Fig 9: Performance Comparison of XGBoost + KNN Variants with PCA and SMOTE

Figure 9 illustrates the performance of XGBoost and KNN combinations with PCA and SMOTE. It shows that applying both PCA and SMOTE together improves accuracy while reducing errors compared to the original model. This highlights the effectiveness of hybrid optimization techniques in yield prediction.

Conclusion

This research successfully implemented and evaluated several machine learning models and simple ensemble techniques for the tasks of crop type classification and crop yield estimation. Through a structured process involving data loading, pre-processing, model training (Random Forest, XGBoost, and KNN), and evaluation, key insights into the predictive capabilities of these algorithms were gained.

For classification of crops, XGBoost model showed the best accuracy with 0.954. The Random Forest model also performed reasonably well with an accuracy of 0.766 while the KNN has the lowest accuracy of 0.488. Simple ensemble methods combining these models did not surpass the individual performance of the XGBoost model in terms of classification, the Random Forest + XGBoost ensemble got an accuracy of 0.816 and the XGBoost + KNN got an accuracy of 0.710. This shows the good independent performance of the XGBoost algorithm, it is on this dataset for the classification problem. For predicting crop yield, the models had different predictions for Mean Absolute Error (MAE) score.

VII. Future Scope

The MAE of Random forest model was 0.0177, XGBoost was 0.0199, KNN was 0.0191. The simple weighted averaging ensemble of Random Forest and XGBoost gave MAE as 0.0042 which indicates that there is a potential improvement in yield prediction using this combination. The XGBoost + KNN ensemble had an MAE of 0.0079.

In summary, the XGBoost model was the most effective single model in the crop classification and yield prediction in this study. While the simple ensemble approaches did not improve classification accuracy more than the best individual model, the Random Forest + XGBoost ensemble had promising results for yield prediction.

Based on these results, in future studies advanced ensemble methods such as stacking or weighted voting could be investigated in order to enhance the accuracy of classifying crops even further, especially by exploiting the high-performance XGBoost. Experimenting with different weighting schemes of yield prediction ensembles is also a potential place for improvement. Additionally, discussing more comprehensive hyper parameter tuning, sophisticated feature engineering and testing models on larger and more diverse datasets, would help to create more robust and generalizable crop prediction systems.

References

1. Jiang J, Zhao Y, Xing F, Zeng X, Zou Q. HISTIF: A new spatiotemporal image fusion method for high-resolution monitoring of crops at the subfield level. IEEE J Sel Top Appl Earth Obs Remote Sens. 2020;13:4607-26.
2. Jiang J, Xing F, Zeng X, Zou Q. Investigating maize yield-related genes in multiple omics interaction network data. IEEE Trans Nanobiosci. 2020;19(1):142-51.
3. Mateo-Sanchis A, Adsuara JE, Piles M, Muñoz-Marí J, Pérez-Suay A, Camps-Valls G. Interpretable long short-term memory networks for crop yield estimation. IEEE Geosci Remote Sens Lett. 2023;20:1-5.
4. Mariammal G, Suruliandi A, Raja SP, Poongothai E. Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers. IEEE Trans Comput Soc Syst. 2021;8(5):1132-42.
5. Jaques LBA, Coradi PC, Lutz E, Teodoro PE, Jaeger DV, Teixeira AL. Nondestructive technology for real-time monitoring and prediction of soybean quality using machine learning for a bulk transport simulation. IEEE Sens J. 2023;23(3):3028-40.
6. Zhang F, Hassanzadeh A, Kikkert J, Pethybridge SJ, Van Aardt J. Evaluation of leaf area index (LAI) of broadacre crops using UAS-based LiDAR point clouds and multispectral imagery. IEEE J Sel Top Appl Earth Obs Remote Sens. 2022;15:4027-44.
7. MirhoseiniNejad SM, Abbasi-Moghadam D, Sharifi A. ConvLSTM-ViT: A deep neural network for crop yield prediction using earth observations and remotely sensed data. IEEE J Sel Top Appl Earth Obs Remote Sens. 2024;17:1-15.
8. Dey T, Bera S, Latua LP, Parua M, Mukherjee A, De D. iCrop: An intelligent crop recommendation system for Agriculture 5.0. IEEE Trans AgriFood Electron. 2024;2(2):587-95.
9. Gao H, Wang C, Wang G, Li Q, Zhu J. A new crop classification method based on the time-varying feature curves of time series dual-polarization Sentinel-1 data sets. IEEE Geosci Remote Sens Lett. 2020;17(7):1183-7.
10. Nejad SMM, Abbasi-Moghadam D, Sharifi A, Farmonov N, Amankulova K, Laszlz M. Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. IEEE J Sel Top Appl Earth Obs Remote Sens. 2023;16:254-66.
11. Chamorro Martinez JA, Oliveira H, Santos JAD, Feitosa RQ. Open set semantic segmentation for multitemporal crop recognition. IEEE Geosci Remote Sens Lett. 2022;19:1-5.
12. Wang X, Xu Y, Wang H, Kang M, Hua J, Wang F-Y. Region-farm crop planning through double deep Q-learning toward sustainable agriculture. IEEE Trans Comput Soc Syst. 2024;11:1-10.
13. Kierdorf J, Roscher R. Reliability scores from saliency map clusters for improved image-based harvest-readiness prediction in cauliflower. IEEE Geosci Remote Sens Lett. 2023;20:1-5.
14. Udutalapally V, Mohanty SP, Pallagani V, Khandelwal V. SCrop: A novel device for sustainable automatic disease prediction, crop selection, and irrigation in Internet-of-Agro-Things for smart agriculture. IEEE Sens J. 2021;21(16):17525-17538.
15. Chang YJ, Lai MH, Wang CH, Huang YS, Lin J. Target-aware yield prediction (TAYP) model used to improve agriculture crop productivity. IEEE Trans Geosci Remote Sens. 2024;62:1-11.
16. Gupta R, Singh A, Choudhary S, Mittal S, Garg R, Sharma P, *et al*. WB-CPI: Weather based crop prediction in India using big data analytics. IEEE Access. 2021;9:137869-85.
17. Badshah A, Alkazemi BY, Din F, Zamli KZ, Haris M. Crop classification and yield prediction using robust machine learning models for agricultural sustainability. IEEE Access. 2024;12:1-15.
18. Sarkar S, Patel A, Banerjee S, Reddy P, Singh R. Crop yield prediction using multimodal meta-transformer and temporal graph neural networks. IEEE Trans AgriFood Electron. 2024;2:1-9.
19. Najjar H, Miranda M, Nuske M, Roscher R, Dengel A. Explainability of sub-field level crop yield prediction using remote sensing. IEEE J Sel Top Appl Earth Obs Remote Sens. 2025;18:1-12.
20. Shafi U, Iqbal N, Mahmood T, Zaidi S, Khan M, Ali S, *et al*. Tackling food insecurity using remote sensing and machine learning-based crop yield prediction. IEEE Access. 2023;11:108640-57.
21. Munaganuri RK, Rao YN. PAMICRM: Improving precision agriculture through multimodal image analysis for crop water requirement estimation using multidomain remote sensing data samples. IEEE Access. 2024;12:52815-36.
22. Sharma P, Dadheech P, Aneja N, Aneja S. Predicting agriculture yields based on machine learning using regression and deep learning. IEEE Access. 2023;11:111255-64.
23. Saleem RM, Khan A, Iqbal S, Farooq M, Aslam M, Javed A, *et al*. Internet of Things based weekly crop pest prediction by using deep neural network. IEEE Access. 2023;11:85900-13.
24. Nigar N, Faisal HM, Umer M, Oki O, Lukose JM. Improving plant disease classification with deep-learning-based prediction model using explainable artificial intelligence. IEEE Access. 2024;12:100005-14.
25. Abidin MAZ, Mahyuddin MN, Zainuri MAA. Optimal efficient energy production by PV module tilt-

orientation prediction without compromising crop-light demands in agrivoltaic systems. IEEE Access. 2023;11:71557-72.
26. Kumar S, Pant M, Nagar A. Forecasting the sugarcane yields based on meteorological data through ensemble learning. IEEE Access. 2024;12:1-12.
27. Shafi U, Ali S, Aslam M, Hussain A, Tariq S, Javed A, *et al*. Embedded AI for wheat yellow rust infection type classification. IEEE Access. 2023;11:23726-38.
28. Kim DW, Jang G, Kim HJ. Development of CNN-based semantic segmentation algorithm for crop classification of Korean major upland crops using NIA AI HUB. IEEE Access. 2025;13:1-12.
29. Erlin A, Yunianta, Wulandhari LA, Desnelita Y, Nasution N, Junadhi. Enhancing rice production prediction in Indonesia using advanced machine learning models. IEEE Access. 2024;12:1-14.
30. Ikram A, Aslam W. Enhancing intercropping yield predictability using optimally driven feedback neural network and loss functions. IEEE Access. 2024;12:162769-87.
31. Dwivedi AK, Singh AK, Singh D, Kumar H. Development of an adaptive linear mixture model for decomposition of mixed pixels to improve crop area estimation using artificial neural network. IEEE Access. 2023;11:5714-23.
32. Mahmud T, Alam M, Razzak M, Islam M, Sultana N, Hossain S, *et al*. An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning. IEEE Access. 2024;12:1-12.
33. Xu X, Patibandla RSM, Arora A, Al-Razgan M, Awwad EM, Nyangaresi VO. An adaptive hybrid (1D-2D) convolution-based ShuffleNetV2 mechanism for irrigation levels prediction in agricultural fields with smart IoTs. IEEE Access. 2024;12:71901-18.
34. Mitra A, Singh R, Patel R, Ahmed F, Das S, Banerjee A, *et al*. Cotton yield prediction: A machine learning approach with field and synthetic data. IEEE Access. 2024;12:101273-88.
35. Krishna R, Prema KV. Constructing and optimizing RNN models to predict fruit rot disease incidence in areca nut crop based on weather parameters. IEEE Access. 2023;11:110582-95.